Statistical mining of predictability of seasonal precipitation over the United States

William K. M. Lau[1], Kyu-Myong Kim[2], Samuel S. P. Shen[3]

Popular Summary

In this paper, we address the fundamental issue of overcoming the so-called "spring predictability barrier", i.e., the dramatic drop in prediction skills from boreal spring to summer, that is endemic in forecasting of seasonal rainfall anomalies over the United States based on El Niño. For this purpose, we seek to maximize the predictive information for seasonal precipitation forecast over the US from a variety of predictor fields, by developing an ensemble canonical correlation (ECC) prediction scheme. The ECC carries out independent forecasts from various predictor fields, e.g., SST, soil moisture and snow cover, and then optimally combine the individual forecasts to produce an ensemble forecast. Using 49 years (1951-1999), we apply the ECC to forecast seasonal rainfall anomalies over the US from SST in five non-overlapping sectors: the tropical Pacific, North Pacific, tropical Atlantic, North Atlantic and the Indian Ocean. Results show that ECC yields a remarkable (10-20%) increase in the baseline prediction skills for all regions of the US and for all seasons compared to traditional statistical prediction schemes. We find that while El Niño provides the bulk of the precipitation prediction skill for the boreal winter in the southern-tier states of the US, the regions of the northern Great Plains and the mid-west are not directly affected by El Nino, but rather from SST signals in the North Pacific in boreal summer. Most importantly, the ECC significantly reduces the spring predictability barrier over the conterminous US, and substantially raises the skill bar for dynamical seasonal rainfall prediction.

[1]NASA/Goddard Space Flight Center
[2]Science Systems and Applications, Inc.
[3]National Research Council

Statistical Mining of Predictability of Seasonal Precipitation over the United States

William K. M. Lau and Kyu-Myong Kim

Climate and Radiation Branch, NASA Goddard Space Flight Center, Greenbelt,

Maryland

Samuel S. P. Shen

Department of Mathematical Sciences, University of Alberta, Edmonton, Canada

Abstract

Results from a new ensemble canonical correlation (ECC) prediction model yield a remarkable (10-20%) increase in baseline prediction skills for seasonal precipitation over the US for all seasons, compared to traditional statistical predictions. While the tropical Pacific, i.e., El Nino, contributes to the largest share of potential predictability in the southern tier States during boreal winter, the North Pacific and the North Atlantic are responsible for enhanced predictability in the northern Great Plains, Midwest and the southwest US during boreal summer. Most importantly, ECC significantly reduces the spring predictability barrier over the conterminous US, thereby raising the skill bar for dynamical predictions.

It is well known that sea surface temperature (SST) in the tropical Pacific associated with the El Nino underpins the enhanced forecasting skill for the United States (US) precipitation during the boreal winter. However the skill drops dramatically in the spring, reaches a minimum in the warm season, and rises steadily from fall to winter (1, 2). The dramatic reduction in forecast skill from winter to summer through the spring season is known as the "spring predictability barrier" which has been endemic in both statistical and dynamical forecasts of El Nino (3). Recently, using singular value decomposition, significant predictability was found from the tropical and extratropical Pacific SST on the warm season precipitation over the upper Great Plains and Atlantic States of the US during El Nino summers (4). The increased predictability is a quantitative validation of earlier findings on the relation between the US precipitation and the tropical and North Pacific SST (5, 6). However, the forecasting skill was still relatively low in summer, even during time of strong SST signal in the tropical Pacific.

It has been suggested that the reduced precipitation predictability in the summer over the US stems from the weaker, and more poleward position of the upper level westerly flow in the northern hemisphere, making it more difficult for tropical SST influence to be transmitted to the US continent (7). As a result, the influence of tropical Pacific SST on the US summertime precipitation diminishes significantly. However, climate variability in other regions, especially the extratropics may begin to have an impact on US precipitation in summer (8). The North Atlantic Oscillation and the North

Pacific Oscillation have strong impressions on the SST over the North Atlantic and the North Pacific, respectively, which can be used as potential predictors for US precipitation. However, due to the vast area and dominant SST variance of the tropical Pacific, the El Nino/Southern Oscillation (ENSO) tends to overshadow any influences from SST in other ocean basins. In addition to SST, other factors such as soil moisture, snow cover and vegetation may influence US precipitation predictability for both summer and winter (9, 10). The ensemble canonical correlation (ECC) prediction model has been developed at the NASA/Goddard Space Flight Center, with the purpose of systematically exploring potential predictability associated with the aforementioned factors. Mathematical details of the ECC scheme and preliminary results have been reported (11). In this article, the ECC prediction model is introduced and results for seasonal US precipitation prediction is presented. By deriving maximal predictive information from SST independently from ocean basins, the ECC yields a substantial increase in the baseline prediction skill for US seasonal precipitation for all seasons, and greatly reduces the spring predictability barrier.

Data and the ECC Method

The precipitation data used for this study are derived from optimal interpolation from over 17,000 stations in the Global Historical Climatological Network Version 2 and the Climate Anomaly Monitoring System for the period 1951-1999 (12). The data cover the global land with a spatial resolution of 2.5 degrees latitude-longitude. In this work, only data over the US continent are used. The SST data are obtained from the US

3

National Center for Environmental Prediction for the same period with a spatial resolution of 2 degrees latitude-longitude (*13*). To reduce small-scale noise, the SST data are further averaged to boxes of 6 degrees longitude and 4 degrees latitude.

The ECC prediction model is based on linear regressions that maximize the correlation between the weighted integral of SST and precipitation fields (*14*). The discretization of the canonical correlation in EOF spectral space for the spatially continuous predictor and predictand fields leads to an equal area-factor correction. The regression error is estimated for each EOF mode (*11*). The inclusion of the error estimation and area-factor represents an improvement of the Canonical Correlation Analysis (CCA) method used for operational forecasts at the US Climate Prediction Center (*2, 15*).

The ECC procedure is described briefly in the following. First, the monthly anomaly data of SST and precipitation are obtained by removing the 49-year climatology and linear trend and normalized by the sample standard deviation at each grid box. Next, the matrix for the correlation eigen-problem is solved in the EOF space to obtain the maximum correlation between the canonical correlation variables of SST and precipitation. The predicted precipitation field, $P(t+\Delta t)$, is then expressed in terms of the canonical variables of $SST(t)$, where t denotes time and Δt is the forecast lead time. To maximally extract precursory signals in global SST, the world ocean is partitioned into non-overlapping sectors, and separated forecasts are made based on different ocean sectors. The ensemble forecast is then obtained at each grid box as a weighted average,

of the individual forecasts similar to the "super-ensemble" technique (*16*). An added benefit of the ECC approach is that by ranking the skills of the individual forecasts at each grid box, it is possible to identify which ocean basins contribute maximally to the precipitation predictability over specific sub-areas of the US continent.

For the hindcasts of a particular year, the EOFs and canonical correlation of *SST(t)* and *P(t+Δt)* are computed using the other 48 years. In this way, 49 hindcasts can be obtained. For evaluating the potential predictability, Δt is zero. The zero-lag "prediction" represents the maximal predictability for precipitation given perfect knowledge of the simultaneous SST field. Alternatively, the zero-lag ECC prediction can be used in conjunction with a two-tier forecast scheme in which the SST is predicted by an ocean model or a coupled model. When $\Delta t > 0$, the ECC method can be used as a stand-alone statistical forecast scheme.

Evaluation of Potential Predictability

To evaluate the forecast skill, a number of precipitation forecast skill scores, including the spatial pattern correlation, Hiedke (2-category) score, and the three-category "hit" score, have been computed (*4*). The results reported here are robust and independent of the choice of skill score. For brevity, only the three-category hit scores are discussed here. For each grid box, the observed precipitation values for a given season in 49 years are sorted in ascending order. Three categories are formed according to the first third (below normal), the middle third (normal), and the last third percentiles

(above normal). If the forecast and observed precipitations are in the same category, the forecast is a "hit". The forecasting skill is the hit rate, which is the number of correct forecasts divided by the total number of years, i.e., 49. For a no-skill random forecast, the expected hit score is 33.33%. For a sample size of 49 years the hit rate of 45% (48%) is significantly different from a random forecast at the 5% (1%) significant level.

As a test of the ECC, the hit scores for US wintertime (DJF) and summertime (JJA) predictions averaged over 49 years, based on SST in individual ocean basins, have been computed. The ocean basins are the tropical Pacific (TPAC, 30°S-30°N), the North Pacific (NPAC, north of 30°N), the tropical Atlantic (TATL, 30°S-30°N), the North Atlantic (NATL, north of 30°N), and the Indian Ocean (IND, north of 30°S). We have also computed the skill score using the global ocean, i.e., all ocean basins. The all-ocean skill score is comparable to that computed from TPAC, because of the dominance of the ENSO signal in an all-ocean SST EOF decomposition. In all the results shown, six dominant EOF modes are used. The skill scores vary only slightly if more EOFs have been used. Fig. 1 shows the DJF forecast results using the SST from TPAC, NPAC, TATL, and NATL. Fig. 1a indicates that TPAC has the overall highest score and the most spatially coherent score pattern, concentrating in the southwest US/Mexico and the southeast US. The NPAC (Fig. 1b) contributes significant scores (>45%) in the west and southwest of US and the Great Lakes and Ohio Valley. The TATL (Fig. 1c) and IND (not shown) appear to have the least skill scores, compared to the other ocean basins while the NATL (Fig. 1d) is responsible for the high hit rates in the Pacific Northwest, northeast

6

and southwest US. The skill score for the boreal summer (JJA) from every ocean basin, as shown in Fig. 2, is much reduced and less organized, with the exception of the NPAC, which appears to produce significant prediction skill in a region stretching from the Gulf Coast of Texas to the northern Great Plains and the Midwest. The skill scores shown in Figs. 1 and 2 are comparable to and slightly better than those of traditional CCA methods (*1*)

To evaluate the influence of each ocean basin on precipitation prediction over different regions of the US, each grid box identifies the ocean basin of "maximal" influence, based on the highest temporal correlation between predicted and observed precipitation in 49 years. Fig. 3 shows the distribution of the "influence function" for US precipitation predictability for all four seasons. During DJF (Fig. 3a), it is clear that the TPAC has the strongest influence across the southern states, spanning the southwest, Mexico, the Gulf Coast, the southeast and the eastern seaboard. The TPAC influence reaches up to the mountain states and central US. The NPAC has the strongest influence in the Ohio Valley and the northwest, while the NATL controls the northeastern seaboard, Northern California, Idaho, and Montana. During MAM (Fig. 3b), the influence of the TPAC reduces substantially, while the NATL gains influence in the northeast and along the East Coast. Other regions appear to have competing, but generally weaker influences (relative to the wintertime) from different ocean basins. The previously noted lower skill score in JJA is also reflected in the rather disorganized pattern of the influence function all over the US (Fig. 3c), with perhaps the exception of

northern Great Plains which has the strongest influence from the North Pacific. The JJA pattern suggests a lack of single dominant SST-related forcing mechanism for US summertime precipitation variability. In SON (Fig. 3d), the dominant influence from the NPAC emerges over the Pacific Northwest, the central mountain and southwest states, and the Northern Great Plains/Midwest region. Elsewhere, the TATL appears to have gained influence relative to the other ocean basins. It is clear from the foregoing results that El Nino effect, through SST in the TPAC, is not always the major contributor to rainfall signal over the US, especially in the northern summer. The ECC forecast will capitalize on the additional SST information from ocean basins besides the TPAC.

The ECC forecast is obtained from each individual ocean-basin forecast by assigning an appropriate weight for each forecast at every grid point. In this article, we will show results for the simplest version of the ECC forecast, which is obtained by assigning a weight of one to the most skillful forecast and zero to the rest, based on the 49-year training period. The result of this ECC is not too different from those based on the super-ensemble approach (16) with forecast weights proportional to the regression coefficient. From a comparison of Fig. 4 and Figs. 1 and 2, it is clear that the ECC forecasts raises the skill score in all regions, relative to the forecasts from individual ocean basins as well as from the global ocean (not shown) regardless of the season. In DJF, the skill score increases substantially in the Pacific Northwest and the Great Lakes/Ohio Valley, most due to the inclusion of SST signal from the NPAC and the NATL (see Fig. 1). In JJA, the areas with significantly greater than random forecast skill

8

at 5% significance level increase substantially, especially in northern tier states and in the southwest. The increased score in JJA forecast is mostly derived from SST signal from the NPAC and NATL. Note that the 49-year mean ECC skill scores for DJF and JJA are generally higher and cover more areas than the previous study (4) that was for ENSO years only.

The increase in skill score by the ECC is very robust and is applicable to all regions and all seasons. This claim is supported by considering six representative regions (shown in Fig. 5) of the US, i.e., the North American Monsoon (NAM) region, the Pacific North West (PNW), West Coast/Mountain States (WC/MS), Northern Great Plain/Midwest (NGP/MW), the Gulf Coast (GC), and the Mid-Atlantic (MA). Fig. 6 shows the 49-year mean skill scores for ECC and those for the five individual basins for three-month mean running throughout the entire annual cycle, averaged over the six regions. In Fig. 6, each abscissa month represents a forecast of a three-month mean precipitation centered at the month. For all regions, regardless of the time of the year, there is a substantial increase, ranging from 10-25%, in the ECC skill score compared to those from individual basins. The increase is most notable in the spring and summer, thus greatly reducing the spring predictability barrier. In regions, such as NAM and the GC, the increase in ECC skill is only modest during the boreal winter, presumably because all the predictable SST signal is due to El Nino, which is already maximally extracted from the tropical Pacific. However in other regions such as WC/MS, PNW, and NGP/MW, the wintertime skill scores are also substantially increased.

Most interesting, the NGP/MW regions show a skill score of approximately 50% (> 33.33% at less than 1% significance level) for both summer and winter. In the one-season lag forecast (not shown), the skill score for this region is actually higher in the summer than in the winter, mainly due to the impact of the North Pacific SST. This result is consistent with the recent findings (*17, 18*), which showed that enhanced summertime precipitation in the northern Great Plains and Midwest may be related to the occurrence of recurrent global monsoon modes which has strong SST signature in the North Pacific.

Potential Application of ECC

Results of the ECC forecast model for US seasonal precipitation prediction have shown a remarkable across-the-board increase in prediction skill for all regions regardless of the time of the year. Further increase in skill scores is achieved by stratifying the data according to phases of major climate events such as ENSO and the North Atlantic Oscillation. It is worth noting that the ECC skill reported here, averaged over 49 forecasts, without stratifying, is comparable to or better than the prediction skill of the previous study for ENSO events (*4*). When the skill scores are stratified according to El Nino and La Nina, results (not shown) indicate additional improvement in forecast skills (>60-70% hit rates) can be achieved in the NGP/MW and NAM Region (*19*). Another significance of the ECC forecasting is its implicit use of the nonlinear interaction among the SSTs over different ocean basins and precipitation over the US. The nonlinearity is reflected in the forecasting results since the ECC forecast from all oceans is far better than the sum of the forecasts from individual ocean basins and the forecast from the

10

entirety of all global oceans. We note that predictability may also be further mined by including soil moisture, snow cover, and other regional data that provide additional information independent of large scale SST. Finally, the ECC forecasts can be applied to other climate subsystems and, in conjunction with further diagnostic or model studies will enable a better understanding of the dynamic links between climate variations and precipitation, not only for the US but also for other continental regions.

References and Notes

1. A. G. Barnston, *J. Clim.* **7**, 1513 (1994).

2. A. G. Barnston, T.M. Smith, *J. Clim.* **9**, 2660 (1996).

3. D. K. Chen, S. E. Zebiak, M. A. Cane, A. J. Busalacchi, *Mon. Wea. Rev.* **125**, 773 (1997).

4. H. Wang, M. Ting, M. Ji, *Geophys. Res. Lett.* **26**, 1341 (1999).

5. M. Ting, M., H. Wang, *J. Clim.* **10**, 1853 (1997).

6. C. F. Ropelewski, M. S. Halpert, *Mon. Wea. Rev.* **114**, 2352 (1986).

7. K. M. Lau, L. Peng, *J. Clim.* **5**, 140 (1992).

8. R. W. Higgins, A. Leetmaa, Y. Xue, A. G. Barnston, *J. Clim.* **13**, 3994 (2000).

9. R. D. Koster, M. J. Suarez, M. Heiser, *J. Hydrometeor.* **1**, 26 (2000).

10. M. J. Fennessy, J. Shukla, *J. Clim.* **12**, 3167 (1999).

11. S. S. P. Shen, K. M. Lau, K. M. Kim, G. Li, NASA/*TM-2001-104567, Vol. 17* (NASA/Goddard Space Flight Center, Maryland, 2001).

12. M. Chen, P. Xie, J. E. Janowiak, P.A. Arkin, *J. Hydrometeor.*, submitted (2001).

13. T. M. Smith, R.W. Reynolds, R.E. Livezey, D.C. Stokes, *J. Clim.* **9**, 1403 (1996).

14. K. -Y. Kim, G. R. North, *J. Clim.* **11**, 3046 (1998).

15. T. P. Barnett, R. Preisendorfer, *Mon. Wea. Rev.* **115**, 1825 (1987).

16. T. N. Krishnamurti, C. M. Kishtawal, D. W. Shin, C. E. Williford, *J. Clim.* **13**, 4217 (2000).

17. K. M. Lau, H. Weng, *CLIVAR Exchanges* **5**, 18 (2000).

18. K. M. Lau, H. Weng, GEWEX *News* **10**, May Issue, 4 (2000).

19. K. C. Mo, J. N. Paegle, *J. Clim.* **13**, 3588 (2000). This paper concluded that the Southwest US summer precipitation is modulated by SST anomalies over the TPAC and NPAC.

20. This work was supported by the NASA Global Model and Data Analysis Program, the US National Research Council, and Canada National Sciences and Engineering Research Council. Dr. P. Xie provided the gridded monthly land precipitation data.

Figure Captions

Fig. 1. Three-category hit score (%) for DJF precipitation prediction derived from SST anomalies from a) Tropical Pacific, b) North Pacific, c) Tropical Atlantic, and d) North Atlantic. A hit score of 33% or less indicates the absence of prediction skill. At 5% and 1% significance levels, the greater than 33% hit scores are approximately 45% and 48% respectively. The area with the hit score greater than or equal to 45% is shaded.

Fig. 2. Same as in Fig. 1, except for JJA.

Fig. 3. The "influence function" on US precipitation by SST from dominant variability in different ocean basins. The color indicates the most important influence from the corresponding ocean basins. For example, the red region is most influenced by the Tropical Pacific: (a) Season DJF, (b) season MAM, (c) season JJA, and (d) SON.

Fig. 4. The spatial distribution of the ECC precipitation skill score over the US for (a) DJF and (b) JJA. The area with the hit score greater than or equal to 45% is shaded.

Fig. 5. Boxes showing geographic locations of the regions as labeled.

Fig. 6. The seasonal cycle of the mean seasonal forecast skill for the selected regions: (a) North American Monsoon, (b) Pacific Northwest, (c) West Coast and Mountain States, (d) Great Plains and Midwest, e) Gulf Coast, and f) Mid- Atlantic coast. The thick solid line indicates the ECC forecast. The forecasts from the five individual ocean basins are as indicated, e.g., the red solid line represents the skill from the TPAC.
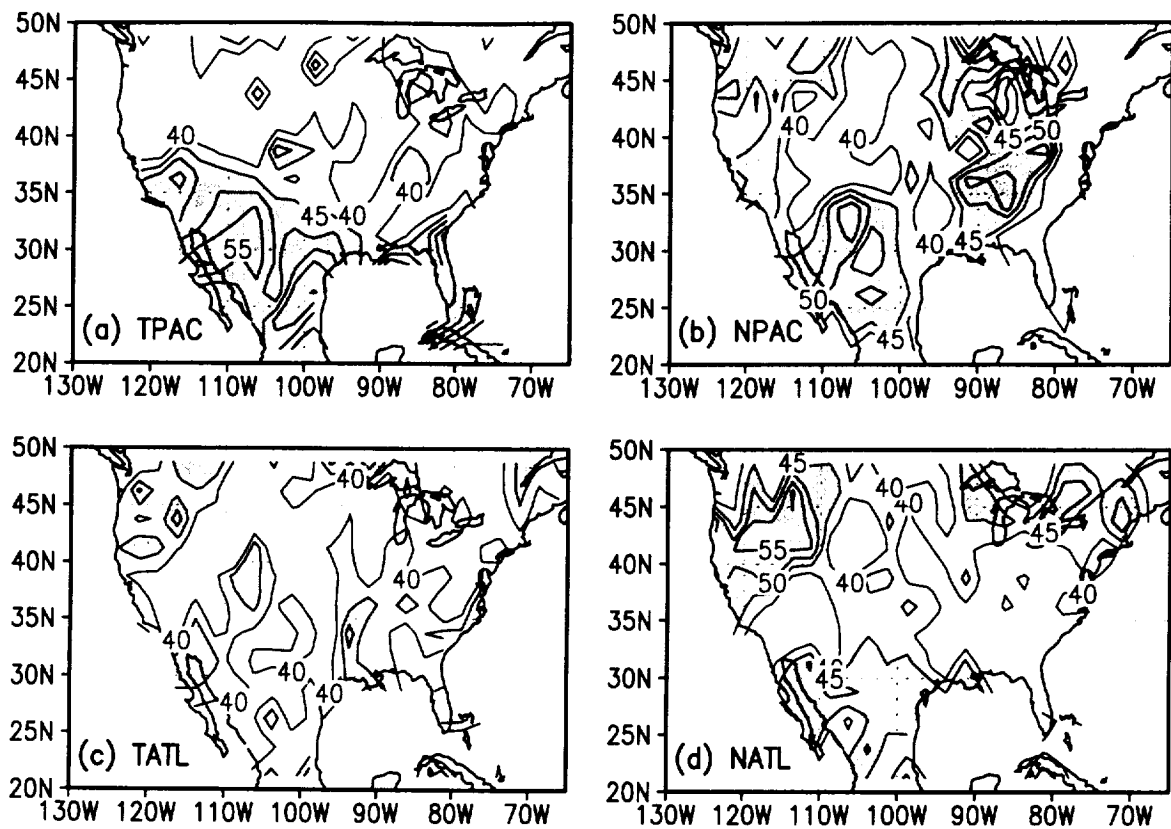
Fig. 1. Three-category hit score (%) for DJF precipitation prediction derived from SST anomalies from a) Tropical Pacific, b) North Pacific, c) Tropical Atlantic, and d) North Atlantic. A hit score of 33% or less indicates the absence of prediction skill. At 5% and 1% significance levels, the greater than 33% hit scores are approximately 45% and 48% respectively. The area with the hit score greater than or equal to 45% is shaded.
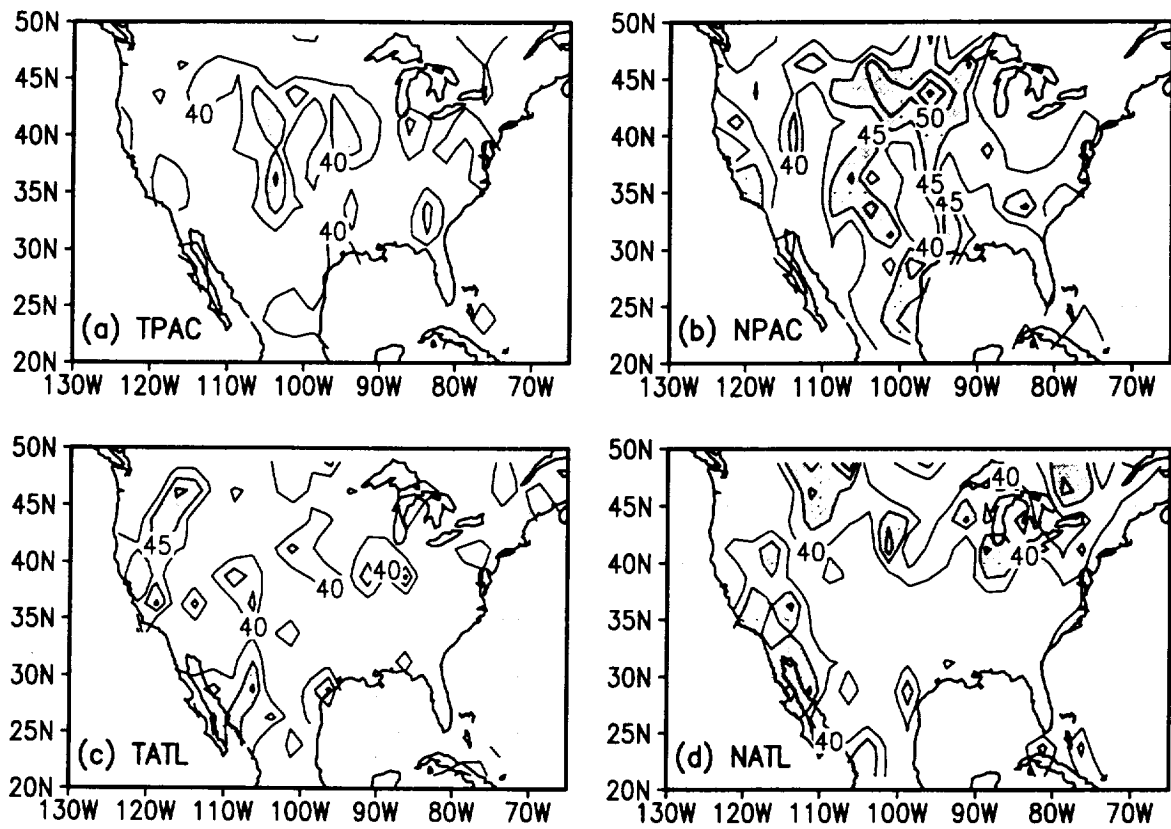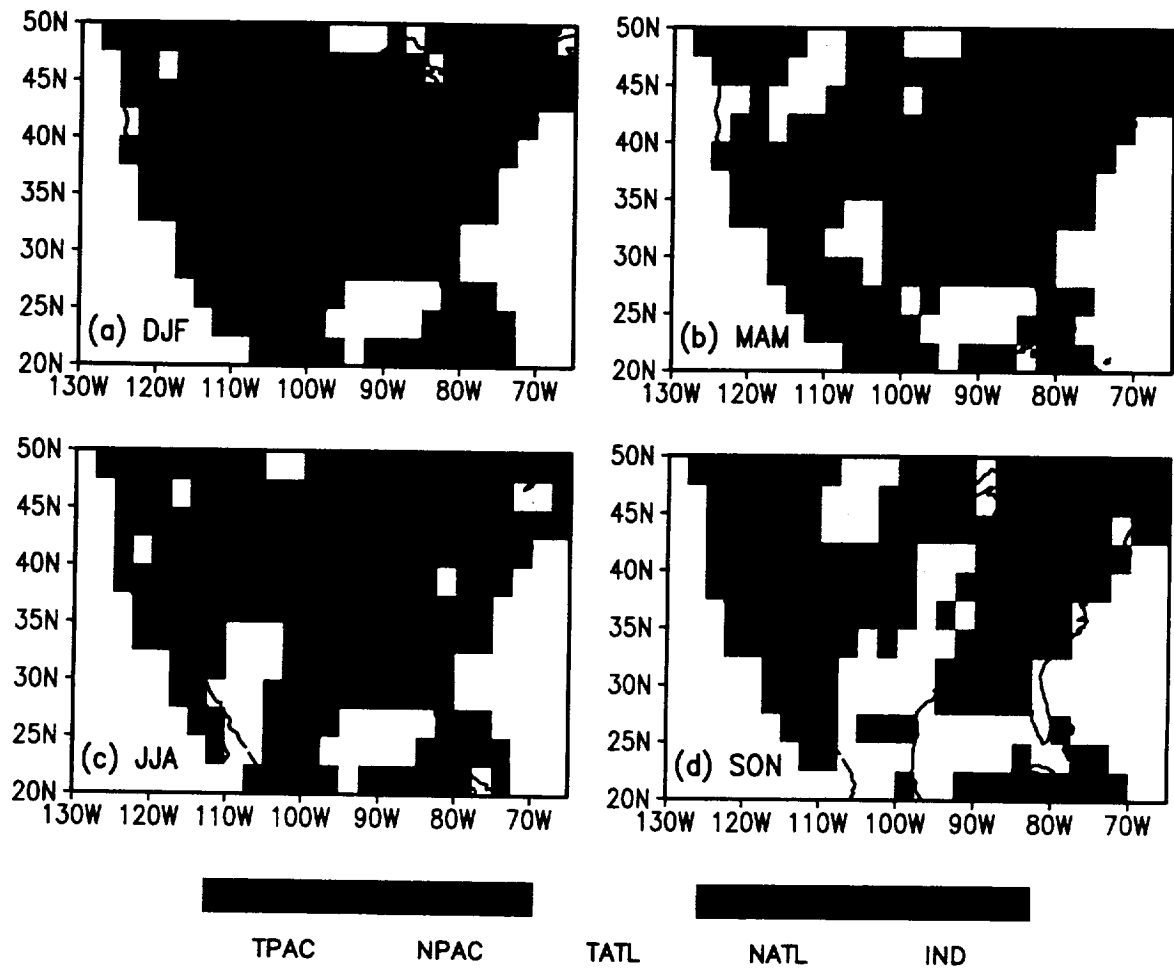
Fig. 2. Same as in Fig. 1, except for JJA.

Fig. 3. The "influence function" on US precipitation by SST from dominant variability in different ocean basins. The color indicates the most important influence from the corresponding ocean basins. For example, the red region is most influenced by the Tropical Pacific: (a) Season DJF, (b) season MAM, (c) season JJA, and (d) SON.
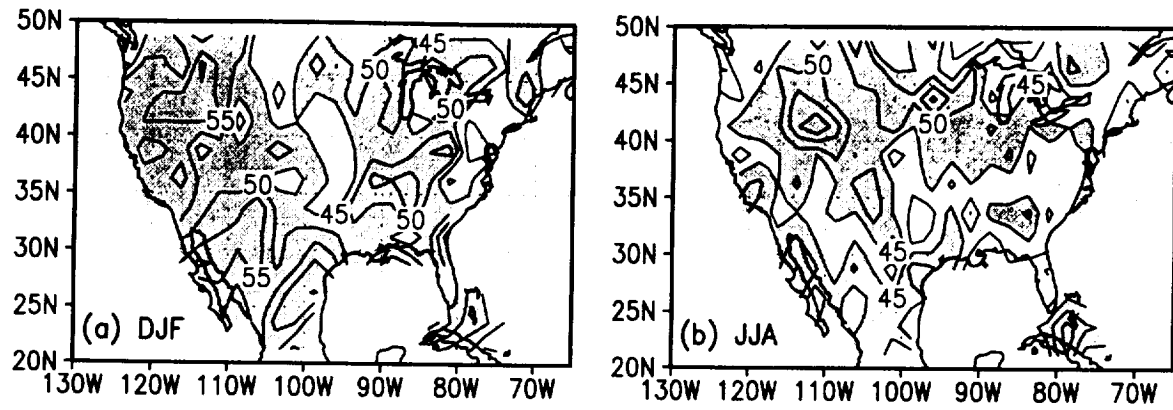
16

Fig. 4. The spatial distribution of the ECC precipitation skill score over the US for (a) DJF and (b) JJA. The area with the hit score greater than or equal to 45% is shaded.
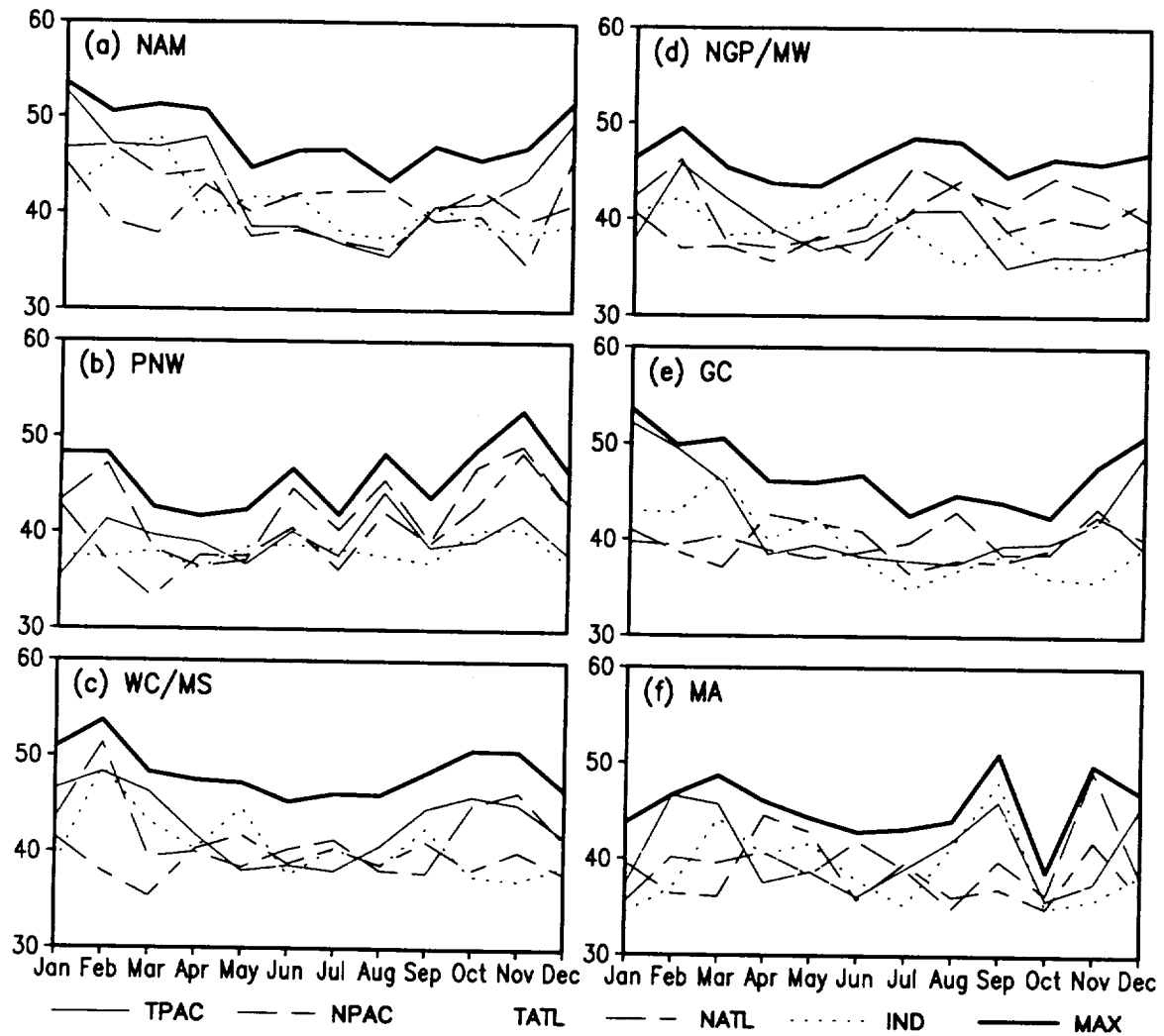
Fig. 6. The seasonal cycle of the mean seasonal forecast skill for the selected regions: (a) North American Monsoon, (b) Pacific Northwest, (c) West Coast and Mountain States, (d) Great Plains and Midwest, e) Gulf Coast, and f) Mid- Atlantic coast. The thick solid line indicates the ECC forecast. The forecasts from the five individual ocean basins are as indicated, e.g., the red solid line represents the skill from the TPAC.

National Aeronautics and Space Administration	NASA Scientific and Technical Document Availability Authorization (DAA)	

The DAA approval process applies to all forms of published NASA Scientific and Technical Information (STI), whether disseminated in print or electronically. It is to be initiated by the responsible NASA Project Officer, Technical Monitor, author or other appropriate NASA official for all presentations, reports, papers, and proceedings that contain NASA STI Explanations are on the back of this form and are presented in greater detail in NPG 2200.2, "Guidelines for Documentation, Approval, and Dissemination of NASA Scientific andTechnical Information.

[X] Original
[ ] Modified

I. DOCUMENT/PROJECT IDENTIFICATION

TITLE

Statistical Mining of Predictability of Seasonal Precipitation over the United States

- AUTHOR (S)

W. Lau (NASA/GSFC)
K.-M. Kim (SSAI)
S. Shen (NRC)

ORIGINATING NASA ORGANIZATION

NASA Goddard Space Flight Center

PERFORMING ORGANIZATION (if different)

CONTRACT/GRANT/INTERAGENCY/PROJECT NUMBER(S)	DOCUMENT NUMBER(S)	DOCUMENT DATE

For presentations, documents, or other STI to be externally published (including through electronic media), enter appropriate information on the intended publication such as name, place, and date of conference, periodical or journal name, or book title and publisher in the next box. These documents must be routed to the NASA Headquarters or Center Export Control Administrator for Approval (see Sections III and VIII).

► **Science**

II. SECURITY CLASSIFICATION

CHECK ONE (One of the five boxes denoting Security Classification must be checked.)

[ ] SECRET   [ ] SECRET RD   [ ] CONFIDENTIAL   [ ] CONFIDENTIAL RD   [X] UNCLASSIFIED

III. AVAILABILITY CATEGORY

[ ] ITAR   [ ] EAR

Export Controlled Document - USML Category /CCL Export Control Classification Number (ECCN)   (Documents marked in this block must have the concurrence/approval of the NASA Headquarters or Export Control Administrator (see Section VIII).

[ ] TRADE SECRET

[ ] SBIR

[ ] COPYRIGHTED

Confidential Commercial Document (check appropriate box at left and indicate below the appropriate limitation and expiration:

[ ] U.S. Goverment agencies and U.S. Government agency contractors only
[ ] NASA contractors and U.S. Government only
[ ] U.S. Government agencies only
[ ] NASA personnel and NASA contractors only
[ ] NASA personnel only
[ ] Available only with the approval of issuing office
[ ] Limited until (date)

[X] **PUBLICLY AVAILABLE**

Publicly available documents must be unclassified, may not be export controlled, may not contain trade secret or confidential commercial data, and should have cleared any applicable patents application process.

IV. DOCUMENT DISCLOSING AN INVENTION

THIS DOCUMENT MAY BE RELEASED ON (date)	NASA HQ OR CENTER PATENT OR INTELLECTUAL PROPERTY COUNSEL SIGNATURE	DATE

V. BLANKET RELEASE (OPTIONAL)

[ ] All documents issued under the following contract/grant/project number may be processed as checked in Sections II and III.

[ ] The blanket release authorization granted on (date)

[ ] is RESCINDED - Future documents must have individual availability authorizations.

[ ] is MODIFIED - Limitations for all documents processed in the STI system under the blanket release should be changed to conform to blocks as checked in Sections II and III.